AD-A284 162

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

...tion is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, ...pleting and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this ...uding this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ..., and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

2. REPORT DATE
07/31/94

3. REPORT TYPE AND DATES COVERED
Annual, 05/01/93-05/31/94

4. TITLE AND SUBTITLE
Local Spatio-Temporal Analysis in Vision Systems: First Annual Report.

5. FUNDING NUMBERS
Grant #
F49620-93-1-0307
3484/ HS
61103D

6. AUTHOR(S)
Wilson S. Geisler, Alan Bovik, Lawrence Cormack, Joydeep Ghosh, David Gilden, Boaz Super

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Center for Vision and Image Sciences
Mezes Hall 330
The University of Texas at Austin
Austin, TX 78712-7789

8. PERFORMING ORGANIZATION REPORT NUMBER
AEOSR-TR- 94 0504

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NL
110 Duncan Avenue, Suite B115
Bolling AFB DC 20332-0001
Attn: Ms. Marilyn McKee
Dr Tangney

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
SEP 07 1994
S F D

11. SUPPLEMENTARY NOTES
Contact person for authors:
CVIS Administrative Assistant, Kirsten Bradbury
e-mail: Bradbury@psyvax.psy.utexas.edu phone: (512) 471-5380

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The aims of this project are to (i) develop a physiologically and psychophysically based model of low-level human visual processing (a key component of which are local frequency coding mechanisms), (ii) develop image models and image-processing methods based upon local frequency coding, (iii) develop algorithms for performing certain complex visual tasks based upon local frequency representations, (iv) develop models of human performance in certain complex tasks based upon our understanding of low-level processing, (v) develop a computational testbed for implementing, evaluating and visualizing the proposed models and algorithms, using a massively parallel computer. Progress has been substantial on all aims. The highlights include, (i) completion of a number of psychophysical and physiological experiments revealing new, systematic and exciting properties of the primate (human and monkey) visual system, (ii) further development of image models that can accurately represent the local frequency structure in complex images, (iii) near completion in the construction of the Texas Active Vision Testbed, (iv) development and testing of several new computer vision algorithms dealing with shape-from-texture, shape-from-stereo, and depth-from-focus, (v) implementation and evaluation of several new models of human visual performance, (vi) evaluation, purchase and installation of a MasPar parallel computer.

14. SUBJECT TERMS

DTIC QUALITY INSPECTED 3

15. NUMBER OF PAGES
30

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
U	U	U	U

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

Local Spatio-temporal Analysis in Vision Systems: First Annual Technical Report

AFOSR F49620-93-1-0307

May 1, 1993 to May 31, 1994

Investigators:

W. S. Geisler
A. C. Bovik
D. G. Albrecht
J. Ghosh
L. K. Cormack
D. L. Gilden
B. J. Super

Submitted July 31, 1994

This is the first progress report of the vision group at the University of Texas under support of AFOSR grant F49620-93-1-0307. We feel that the first year has been very successful and that we are well on the road toward achieving the major aims of the proposal. Before launching into the technical details, we would like to mention several other developments related to the vision group and its AFOSR support.

(a) The AFOSR grant served as a key factor in the formation of the interdisciplinary Center for Vision and Image Sciences (CVIS) at the University of Texas. The Center was awarded renovated space and some additional privileges within the University of Texas system.

(b) The AFOSR grant served as a key factor in the acquisition of additional research space and some equipment from the Departments of Psychology and Electrical and Computer Engineering.

(c) The AFOSR grant has fostered substantially increased interactions between faculty and students in the Department of Psychology, the Department of Electrical and Computer Engineering, and the Biomedical Engineering Program. The nature and depth of some of these interactions is apparent in this progress report (see later).

(d) Many of the investigators supported by the AFOSR grant have played a substantial role (under Bovik's leadership) in bringing into existence a major new annual conference devoted to vision research -- the International Conference on Image Processing (ICIP) -- which is under the IEEE umbrella. The first ICIP meeting (which received over 1000 submitted papers) will be held in Austin in November, 1994.

The original proposal contained five major aims.

Aim 1: *To develop a mathematical model of the initial stages of visual processing (the front-end mechanisms), based upon a wide range of physiological and psychophysical data.*

Aim 2: *To develop new methods and models of local frequency coding.*

Aim 3: *To develop new mathematical models and computer-vision algorithms for performing complex visual tasks that are based upon local frequency coding representations.*

Aim 4: *To develop models for human performance in complex visual tasks that build upon current understanding of the front-end mechanisms.*

Aim 5: *To develop a computational testbed for implementing, comparing, integrating and visualizing the different models and modules developed during the project, using a massively parallel machine and graphics workstation front-end.*

The progress report is organized in an outline fashion. Each bold roman numeral at the top level indicates one of the major aims. Each bold letter at the second level indicates the contributions (toward the aim) from a given investigator's laboratory. Each bold arabic numeral at the third level indicates a specific research project. The names in parentheses following the title of the specific project are the investig and students directly involved in the project. (Also note that each investig s general objectives are briefly summarized the first time that research in that investigator's laboratory is described.) Publications, submitted papers, conference presentations and technical reports are listed at the end of the progress report. Copies of relevant written material are appended. (References marked with an asterisk in the text are contained in the appendix.)

I. Aim 1: *To develop a mathematical model of the initial stages of visual processing (the front-end mechanisms), based upon a wide range of physiological and psychophysical data.*

A. *Albrecht's Laboratory*

The broad objectives of Albrecht's research are:

(a) To characterize the spatio-temporal response properties of neurons in the visual pathway in cats and primates.

(b) To develop a quantitative model of visual neuron responses that includes both linear and nonlinear mechanisms.

(c) To relate the response properties of visual neurons to behavioral performance.

Albrecht's laboratory is designed to measure the responses of individual neurons within the visual systems of monkeys and cats, while stimulus domains of interest are explored in a systematic and quantitative fashion. Over the past decade, Albrecht has described how neurons in the visual cortex respond along several fundamental stimulus dimensions: spatial position, spatial frequency, temporal frequency, spatial contrast, direction of motion, etc. In the process, a new model of the visual cortex has gradually developed. This formal mathematical model builds upon the established principle of linear spatio-temporal filtering by incorporating nonlinearities that were initially revealed through measurements of the contrast response function; these nonlinearities have subsequently been verified through a variety of different measurements in several different laboratories (outside of UT). The model contains four major components: contrast gain, linear filter, expansive exponent, internal noise. This combination of linear and nonlinear components, arranged in a particular sequence, produces several unexpected consequences with respect to how visual cortex neurons transmit, extract and encode information about objects in the environment.

During the last year, with partial support from the AFOSR, Albrecht's laboratory performed a thorough analysis of the noise characteristics of visual cortex neurons and how the noise/variance affects detection, discrimination and identification performance. Albrecht also made substantial progress towards

completing a project which describes the influence of the contrast gain control system on both the phase and the amplitude of the spatio-temporal transfer function.

1. *Measuring and characterizing the noise in the responses of visual cortex neurons (Albrecht, Geisler).*

The performance of the visual system, like any other system, is limited by the variability within the system. Several lines of evidence suggest that the following simple rule might be adequate to describe the noise characteristics of visual cortex neurons: the variance is proportional to the mean, referred to here as the VPM hypothesis. This is neither an obvious hypothesis, nor is it *a priori* a likely hypothesis; this behavior would not be expected on the basis of simple summation of excitatory and inhibitory inputs, given that the variance of the difference of two inputs is as large as the sum of the two. The aim of this study was to perform a rigorous and systematic test of the VPM hypothesis by analyzing both the mean and the variance of the responses of a large sample of neurons, as function of several fundamental stimulus dimensions: spatial position, spatial frequency, spatial contrast, temporal frequency, and direction of stimulus motion. Several different mathematical functions were fitted to the variability data of each neuron using maximum-likelihood methods, so that the adequacy of the fits could be quantitatively compared. Albrecht and Geisler found that the simple one-parameter proportionality rule could account for more than 90% of the trends in the variability data. Further, they found that other functions, such as a two-parameter power function, did not provide a significantly better fit. The simplicity of this rule has a number of pragmatic consequences. For example: (i) Any model that can predict the mean response of cortical cells will be capable of predicting the variance, up to a proportionality constant; (ii) Noise characteristics that accrue from earlier levels in the system can be essentially ignored, regardless of origin; (iii) Constraints are placed on the types of neural computations the brain could utilize at subsequent higher levels of analysis.

All of the data for this project have now been collected and analyzed; a formal research report is being prepared for submission to *Vision Research* [23].

2. *Identification performance of neurons in the primate visual cortex (Albrecht, Geisler).*

Neurons in the visual cortex are selective along a number of stimulus dimensions (e.g. spatial frequency, orientation, position, motion). While this selectivity may be an important, and perhaps even necessary, first step in the process of image identification, selectivity alone is not sufficient. The principle of univariance summarizes the fact that a linear filter cannot uniquely identify image attributes, because an equivalent response can be evoked by a wide range of nonequivalent stimuli (e.g. within a single linear photoreceptor, it is impossible to identify either the wavelength or intensity). Albrecht and Geisler suspected that the contrast gain nonlinearity and the expansive exponent nonlinearity would have a powerful effect on the identification performance of cortical cells. To address these issues we developed a procedure for defining, quantifying and measuring identification

performance in sensory neurons. They then applied this new procedure to measure how accurately primate visual cortex neurons could identify the spatial frequency, the contrast, the direction of motion, etc., based upon the information contained within a single 200 millisecond fixation interval.

The basic approach is to measure the mean and the variance of the response to stimuli which vary along the dimensions of interest and then to use Bayes formula to express the probability of occurrence along these same dimensions, given that a response of some specific magnitude had occurred. In general, for the dimensions measured, they find that when a cortical neuron responds at or near its maximum rate of firing, within an interval of time comparable to a normal fixation, the multidimensional stimulus probability density function is primarily restricted to a very narrow subset of potential stimuli and that certainty along any given dimension is little affected by additional dimensions of stimulus uncertainty. Consider the performance of a representative cell along the dimension of spatial frequency, with the following average properties: 1.2 octave spatial bandwidth at half the maximum response, essentially no maintained activity, variance proportional to the mean, maximum saturated response at 50 spikes/second. When this cell produces 10 (ten) action potentials during a 200 millisecond interval, a later brain mechanism can know, with 95% certainty, that the stimulus was from a band of frequencies slightly greater than 0.8 octaves, even if there is total uncertainty along other stimulus dimensions (such as orientation, contrast, etc.).

This project is complete. A formal research report has been prepared and sent to colleagues at other Institutions for their review, prior to submission to *Science* [29*].

3. *The effect of contrast on the phase transfer function of visual cortex neurons (Albrecht).*

The responses of neurons in the visual cortex of monkeys and cats have been measured as a function of the spatial frequency and the temporal frequency of drifting sinewave grating patterns. These measurements result in a spatio-temporal transfer function. The transfer function provides a quantitative and systematic method for characterizing and describing some of the basic receptive field properties of visual cortex neurons. Further, comparisons of the measured responses to the those expected from a linear system are generally informative. For simple cells in the visual cortex, the phase of the response to drifting sinewave gratings can be affected by four different receptive field properties: the spatial location, the temporal latency, the shape/symmetry of the spatial receptive field, and the shape/symmetry of the temporal receptive field. These four components combine in a very simple fashion, such that the spatio-temporal phase transfer function can be described using linear equations in four parameters; the four parameters provide a quantitative metric for indexing the four receptive field properties.

There have been several brief reports indicating that the phase of the response can also be affected by the contrast: as contrast increases, the phase advances.

Albrecht suspected that this phase advance might be related to the contrast gain nonlinearity; there is some evidence consistent with this view. In this study, he systematically explored the basic effect of contrast on the phase transfer function. The results provide a general qualitative and quantitative description of the overall magnitude and dynamics of the contrast phase relationship. The four basic receptive field properties were estimated at varying levels of contrast; while the spatial properties (position and shape) were largely unaffected by contrast, both of the temporal properties were affected. Both the latency and the shape of the temporal receptive field shifted as a function of contrast. Further, the magnitude of the phase shift was tied to the overall level of stimulating contrast and not the overall level of the neuron's response. Thus, for example, the phase advance induced by a nonoptimal stimulus was similar to the phase advance induced by an optimal stimulus. This fact is consistent with the hypothesis that the phase advance is related to the contrast gain nonlinearity.

All of the data for this project has been collected and analyzed; a formal written research report is nearing completion, and will be submitted to *Visual Neuroscience* [22].

B. *Cormack's Laboratory*

The broad objectives of the research effort in Cormack's laboratory are as follows:

(a) To determine the relevant spatio-temporal contrast information for the sensory/perceptual systems involved in stereopsis.

(b) To determine the relevant spatio-temporal contrast information for the sensory/motor systems involved in vergence eye movements (pending equipment purchase).

(c) To determine the manner in which the spatial contrast information, having been processed in parallel by low-level quasi-linear filters selective for spatial frequency, orientation, etc., is recombined to yield the large disparity range and fine resolution that is observed in behavioral experiments.

(d) To implement the processing algorithms thought to be used by the human visual system into testable, computational models of stereopsis.

Progress on objectives (a) and (b) is described here; progress on objectives (c) and (d) is described under Aim 4.

During the past year, both technical and experimental progress has been made toward the above stated objectives in Cormack's laboratory. Cormack's laboratory is relatively new, so much of the progress has been in the areas of equipment construction, equipment calibration and software development in preparation for the studies outlined in the initial proposal. Also, a number of psychophysical studies have been completed or are in various stages of continuation.

Cormack and Mr. Ramakrishnan (of Computer Engineering) have developed software that enables us to 1) calibrate and linearize the gray scales of the stimulus display monitors, 2) display dynamic random dot displays and/or arbitrary animated sequences computed off-line and stored on disk, 3) display pairs of large stereoscopic images loaded from disk in both a two-alternative forced-choice or yes-no psychophysical format.

Regarding 3) above, Cormack and Chen (from Bovik's laboratory in Computer Engineering) are collaborating on a series of experiments (see later) and have developed the means to import/export images between Cormack's and Bovik's laboratories. This will allow us to use identical stimuli to test both the human observers in Cormack's lab and Chen's computational model in Bovik's laboratory.

1. *Spatio-temporal integration regions in binocular vision (Cormack).*

This study, which has been accepted for publication in *Vision Research* [7*], concerns the spatio-temporal region over which the binocular visual system can integrate information. While not completed under the direct auspices of this grant, much of the research in Cormack's laboratory under the grant will follow from this study. Cormack found that the binocular system behaved as though it could integrate a fixed amount of information, regardless of its distribution in space and time. This surprising result bespeaks the flexibility of biological visual systems.

2. *Vertical and horizontal contrast information in binocular vision (Cormack).*

This study, which will be submitted for publication pending some comments on the manuscript from colleagues [28*], investigated the relative ability of the visual system to utilize horizontal and vertical contrast information in order to combine the information from the two eyes in a meaningful fashion. In theory, the brain must use information from at least the two principle orientations (horizontal and vertical) in order to bring the two eyes' images into "register" even though only horizontal displacements (calculated from vertical edges) are eventually used for computations of relative depth in a scene. Cormack found that the binocular combination of horizontal edges is accessible psychophysically via his correlation detection paradigm; thresholds for horizontal stimuli are comparable to those for vertical stimuli. However, he found that the internal noise distributions for the horizontal stimuli are almost an octave broader, possibly indicating a more restricted spatial integration area along the vertical meridian.

3. *Sampling efficiency in dynamic random-dot stereograms (Cormack)*

This study concerns sampling efficiency in dynamic random dot displays of various densities. Human observers were required to detect the presence of interocular correlation in the these displays. Cormack found that as element density decreased, human performance remained constant over a very broad range, despite the decreasing information content. This is an indication that

typical, dense stereoscopic images are undersampled by the binocular visual system. Moreover, the density at which performance begins to deteriorate provides us with an estimate of the density with which the binocular visual system does sample visual information. Some relevant computational modeling is currently being completed, and a manuscript is being prepared for submission to *Vision Research* [26].

4. *Processing asymmetries in stereopsis (Cormack)*

In this study Cormack is measuring processing asymmetries in the disparity domain. While one of the original theories about the encoding of stereopsis employed separate mechanisms for crossed (near) and uncrossed (far) disparities, more contemporary theories and models generally employ a continuum of 'generic' disparity tuned channels. These contemporary theories predict that there should be no systematic differences in processing of crossed and uncrossed disparities. Yet, Cormack is finding clear quantitative and qualitative differences between the two types of disparity. Qualitatively, he finds that with brief presentation times naive subjects have an extremely difficult time perceiving depth in displays comprising uncrossed disparity. However, depth is readily perceived in displays comprising crossed disparities. Quantitatively, Cormack finds that the processing time for crossed disparities is shorter than for uncrossed disparities; reaction times in 2 alternative forced-choice tasks are consistently faster for crossed disparities. These data will force us to rethink the manner in which disparity is encoded by the human visual system. This study will be submitted for presentation at the 1995 ARVO meeting, and is being prepared for submission to *Perception and Psychophysics* [24].

C. *Geisler's Laboratory*

The broad objectives of Geisler's research are as follows:

(a) To develop a general model of human visual discrimination that can predict performance for a wide range of spatial patterns and adaptation/lighting conditions.

(b) To use current and emerging knowledge of early visual processing, e.g., the discrimination models of objective (a), as a basis for rigorous study of higher-level perceptual and cognitive processing.

(c) To explore computational theories for selected visual tasks in order to better understand the design of the human visual system and to support the development of practical applications in computer vision.

Progress on objective (a) is described here. Progress on objectives (b) and (c) is described under Aims 3 and 4, respectively.

For a number of years Geisler's laboratory has been conducting a parametric examination of how the mechanisms of light and dark adaptation affect spatial pattern vision. The empirical aim of the projects has been to measure amplitude

sensitivity functions (ASFs) for Gabor targets (Gaussian-damped sinewave gratings) under various states of light and dark adaptation. (Note that the ASF is a plot of amplitude sensitivity as a function of target spatial frequency.) The theoretical aim of the projects is to use the data as a basis for a general model of human pattern detection that is applicable under a wide range of adaptation conditions. The experimental work was begun under support from NIH, but has come to fruition during the last year under partial support from the AFOSR grant. The theoretical work has been more fully supported by AFOSR.

1. *Effects of light and dark adaptation on spatial vision (Hahn, Geisler)*

One major study involved measuring (for the first time) ASFs during cone dark adaptation following exposure to very intense adaptation fields (full bleaching fields) and on steady adapting background fields of various intensities. The results are very systematic and enlightening. First of all, Hahn and Geisler found that the shapes of the ASFs measured during dark adaptation were shape invariant (on a log threshold axis). This is a very simple rule which would seem to imply that relative detection threshold across different spatial patterns does not change during cone dark adaptation. Second, Hahn and Geisler found (in agreement with earlier work) that the shapes of the ASFs *do* change systematically as a function of the intensity of the steady adapting background. As is well known, these results imply that the relative detection sensitivity for fine spatial patterns increases as background intensity level increases. This study demonstrates a fundamental dichotomy between bleaching adaptation and background adaptation. The results are playing an important role in the development of a general model of pattern detection and should be of practical value in predicting visibility under conditions of light and dark adaptation. This study is described in Hahn and Geisler [12*].

2. *Spatial vision under transient light conditions (Kortum, Geisler)*

Another major study involved measuring (also for the first time) ASFs in the so-called "probe-flash" paradigm. Specifically, thresholds for Gabor patterns were measured on flashed backgrounds of various intensities, over a wide range of target (i.e., probe) spatial frequencies. These probe-flash curves (plots of threshold vs. flashed-background intensity) were measured in the dark-adapted eye and in the presence of steady adapting backgrounds of various intensities. There were two major goals of the study: one was to obtain parametric data on pattern visibility under a wide range of lighting conditions; the other was to measure the strengths of the subtractive and multiplicative components of light adaptation across target spatial frequency. One major finding was that probe-flash curves change shape systematically with target spatial frequency, suggesting that the different spatial-frequency channels have different luminance nonlinearities. Another major finding was that the strength of multiplicative and subtractive adaptation did not vary greatly across target spatial frequency. These results are also playing a key role in the development of a general model of pattern detection. An immediate practical outcome of this study was a set of simple descriptive formulas that allows approximate prediction of sinewave target visibility (ASF's)

under a wide range of transient lighting conditions. The results are described in Kortum and Geisler [14].

3. *Model of pattern detection (Geisler, Kortum).*

An on-going project in Geisler's laboratory is the development of a quantitative model of pattern detection that can predict detection thresholds under steady state and transient background luminance conditions, and during dark and light adaptation. Most of the components in the model were outlined in the original proposal (and in Geisler, 1994) and will not be summarized here except to point out that the components are based upon current physiological and psychophysical results. Many of the parameters in the model are set directly from anatomical and physiological measurements. The remaining parameters are estimated from certain fundamental psychophysical data. However, the exact model has several nonlinear stages making the calculation of the predictions rather slow (when trying to estimate free parameters). To deal with this problem, Geisler has developed a small-perturbation approximation of the model that allows fast calculations of predictions for detection thresholds under steady background conditions. Geisler and Kortum have used this small perturbation approximation to estimate most of the free parameters. The model was found to accurately predict increment threshold functions for Gabor targets of various spatial frequencies, and to yield plausible parameter estimates. The next step will be to generate predictions for various transient conditions using the exact model. The results to date are briefly described in an invited presentation which was given at the ARVO 1994 meeting. The text and figures from that presentation are given in Geisler [51*].

4. *Model of pattern discrimination (Geisler, Seay).*

Development of the pattern detection model is one stage in the development of a general model of pattern discrimination. Geisler and Seay have been implementing the full pattern discrimination model in Wavax (a "home-grown" modeling environment) and testing the predictions of certain components of the model in demonstrations which have been implemented in Microsoft Excel and Wavax. The demonstration programs have largely been directed at exploring the psychophysical predictions of the Contrast-Gain Exponent (CGE) model of cortical cell responses proposed by Albrecht and Geisler; the CGE model is a key component in the full discrimination model. To date, Geisler and Seay have demonstrated that the CGE model predicts appropriate contrast discrimination functions, and predicts the psychophysical dissociation that has been observed between contrast discrimination and shape (or position) discrimination.

II. Aim 2: *Develop new methods and models of local frequency coding.*

A. *Bovik's Laboratory*

Research being conducted in the Laboratory for Vision Systems (LVS; A.C. Bovik, Director) in the Center for Vision and Image Sciences (CVIS), both in-house and

in conjunction with other faculty and scientists in CVIS, are directed toward two main complementary objectives:

(a) To development Modulation Models and Multiband Modulation Energy Operators for image modeling and analysis.

(b) To development Multiresolution, Foveated, Computational Visual Processing for Active Vision.

Progress on objective (a) will be discussed here; progress on objective (b) will be described under Aim 3.

A broad objective of Bovik's research program is the further development and application of recently introduced, potentially very powerful, general multiresolution Image Modulation Models of contrast and phase structures in image data. The new Image Modulation Models capture information-bearing variations in images as amplitude- and frequency-modulated (AM & FM) sinusoidal functions, d especially, as sums of such functions. The approach broadly generalizes the sinusoidal models commonly used both in studies of biological visual perception (e.g., the spatial-frequency channel models, and the so-called "energy" models of spatio-temporal vision), and also in engineering analysis of digitized images (viz: the Fast Fourier Transform, which decomposes images into sums of sinusoidal functions, but only on a *global* (nonlocal) basis). The new image model has particular potential for analyzing nonstationary image data, and (when coupled with multiband/wavelet decompositions) for the computation of symbolic descriptions of space-varying (nonstationary) modulated image structure.

Complementing the development of the Image Modulation Models are efforts toward extending and analyzing new, conceptually simple, Modulation Energy Operators which supply a powerful framework for the computational demodulation/decoding of modulated image information in machine vision applications, and potentially as a new model for explaining image demodulation in biological visual science. Bovik and collaborators are applying these paradigms to important computational vision applications, including multiband modulation and energy-based demodulation models for coding and representation of image information.

Another general application of these paradigms is to problems in 3-D machine vision. This involves the use of newly-developed 3-D to 2-D nonstationary, spatially localized, surface-to-image Frequency Projection Models. Bovik and collaborators are currently applying the Frequency Projection Models for the computation of three-dimensional scene structure from one or more two-dimensional images. The overall application paradigm embraces and unifies concepts of shape-from-texture, multiband stereopsis, and active visual sensing through the use of local spatial frequency information captured using multiband, multiscale (wavelet) image decompositions and Modulation Energy Operators. Of particular emphasis is our work in Shape-from-Texture and Stereopsis.

1. ***Modulation Models and Multiband Modulation Energy Operators for Image Modeling (Bovik, Havlicek, Pattichis).***

In the recent past, Bovik's laboratory has developed single-component AM-FM models for image processing and analysis, with great success for images that satisfy such a model. However, most images contain multiple superimposed components, necessitating the development of a more general and powerful multicomponent model. Thus, Bovik's group is now developing *Multicomponent* Modulation Energy Operators that are capable of separating and demodulating multiple superimposed modulation components in image data. These multicomponent operators make use of multiband (wavelet) decompositions of the image data into spectrally separated channels that are individually processed at each image location, and then spatially aggregated. Specifically, because an image may contain multiple tracks (which may increase or decrease in amplitude or vanish altogether, or which may merge into fewer components or split into more) a Kalman-filter-based strategy has been developed to *track* the individual AM-FM components across the image domain. The significant results in this effort are just beginning to emerge; however, one Ph.D. student involved in the project has been admitted to candidacy, and two conference papers have been accepted to appear [63, 64]. Several journal papers are in the near-submission stage [24, 32, 34].

2. *Modulation Models and Multiband Modulation Energy Operators in Shape-from-texture (Bovik, Super).*

Bovik and Super have continued development of new theories for practical multiband Shape-from-Texture. This maturing work has demonstrated the feasibility of acquiring truly accurate surface shape/orientation information from a single camera view, using textural information as the basis for computation. The approach taken has combined multiscale wavelet-like image decompositions, combined with Image Modulation Models as the image representation, and Modulation Energy Operators as the processing tools used to extract the necessary projected image data. The use of Frequency Projection Models allows the relationship between surface frequencies, surface shape, and image frequencies to be computed from the multi-channel energy operator outputs. Bovik and Super believe that the results obtained demonstrate unprecedented accuracy, generality, and flexibility relative to prior shape-from-texture paradigms. This work has recently been accepted to appear as two refereed journal publications [16*, 17*].

3. *Modulation Models and Multiband Modulation Energy Operators in Stereopsis* (Bovik, Chen, Cormack).

Bovik, Chen and Cormack have developed new theories for practical multiband stereopsis. Specifically, a multichannel Gabor filter processing paradigm has been designed which also combines Image Modulation Models as the image model, Modulation Energy Operators as the feature-extracting apparatus, and Frequency Projection Models in order to model 3-D - to - 2-D image projections and, in turn, to compute the 2-D - to - 3-D depth computation. Because the perspective projections of 3-D objects can be described as a phase shift, the

correspondence problem is solved by demodulating the outputs of the Gabor channels, and using the local phase information across channels as the matching primitives. A dense, highly accurate depth map is obtained, comparable to any existing stereo algorithm in experiments conducted so far. One Ph.D. student has been admitted to candidacy on this project. This recently developing work will be presented at several conferences [57, 60].

III. Aim 3: *Develop new mathematical models and computer-vision algorithms for performing complex visual tasks that are based upon local frequency coding representations.*

A. *Bovik's Laboratory*

A major goal in Bovik's laboratory is to provide a general platform for demonstrating and evaluating the efficacy of the various paradigms/models developed in both the computational and the physiological/psychophysical components of the proposed research. Specifically, the objective is to complete the construction of a state-of-the-art active vision system with computer-controlled vergence, baseline adjustment, pan, tilt, and focus control. This innovative platform is being designed to naturally incorporate *multiband processing* in a *multiresolution, foveated processing paradigm*. Creating a graded resolution hierarchy emanating from a central fovea makes possible complexity-graded (vergent) stereoscopic processing. Thus, the burden of obtaining detailed, high-resolution scene information is being placed on the design of *fixation, vergence and focusing* strategies. In this way, dense 3-D scene representations are to be obtained by introducing multiple fixation points, making computation vastly more efficient. The critical elements of this ongoing research are the development of multiband foveal structures which balance the need for high resolution information (for recovery algorithms) with the need for a reduced volume of data. This requires the development of active vision control strategies for directing a pair of foveal image sensors to obtain depth and shape information, and the development of dynamic control of the storage of the 3-D reconstructed surface representation.

This long term project (in the sense that much equipment acquisition and system construction is involved at the outset) has the following subgoals:

(*a*) *Platform construction*. One subgoal is to construct an active vision system that is fully software-controllable. The aim is to obtain a system that combines hardware and software protocols for computer feedback-controlled variable baseline vergent stereo, and for lens parameter control (including zoom, depth-from-focus, and dynamic aperture control). The objective is a *highly reliable* platform, with the processing flexibility to allow for multiresolution and foveated image data processing (the breakdown tendency of systems at other laboratories is much higher).

(*b*) *Foveation Strategies*. Defining an effective multiresolution foveation protocol presents different difficulties in theory and in application. With current hardware, theory can at best be coarsely approximated by implementation in most

cases, if reasonable computation time is to be maintained. Within a theoretical framework that may be modified somewhat in application, Bovik and collaborators are implementing processing strategies that emulate a focal plane array or sensing arrangement having a nonuniform sampling pattern. In practice, a uniform, dense array is being used, with the processing proceeding on the data in a hierarchical nonuniform fashion, using a multiband formulation.

(*c*) *Focusing and Fixation Strategies*. Another subgoal is to develop theory and software strategies for automated focusing and fixation strategies that will meld smoothly with a multiresolution, foveated processing framework. The objective is to implement active focus control as an integral component of the system. Active focus control (*not* "autofocus") can be used to directly estimate the depth of points or regions in a scene. By adjusting the lens focus so the maximum "sharpness," near the fovea, is obtained, the lens position can be directly converted into a depth measurement for that point. This technique does not rely on correspondence of points between the cameras, so that it is possible to employ this method as an *independent depth* estimate, thus adding redundancy to the depth computation problem. Various image sharpness criteria (to measure the degree of local or global image focus) are being explored.

Camera fixation is a difficult and often application-specific problem. However, the goal here is the development of a generic camera pointing system that can be adapted to applications, but which also demonstrates the possibility of a fixating system operating without a limiting application in mind. Thus, in the absence of directive intelligence, the active vision system is intended to operate in a freeform fixation mode most interested in either the most rapidly changing area of the scene (motion-cued) or the surface of the object most recently analyzed. In the first case, a region of interest is defined; in the second, a semirandom fixation strategy will attempt to explore the entirety of the surface with particular emphasis on high-information features.

(*d*) *Vergence Strategies*. The proposed foveation strategy is ideally defined for computing depth from a pair of vergent (non-parallel) cameras, since the tradeoff between matching complexity (highest near the periphery in a vergent system) and depth resolution (lowest near the periphery in a foveated system) is to be made explicit in a natural way. This geometry, which is probably not advisable for an inactive static, single-resolution stereo system (which is why vergent computational stereo systems have received only a small amount of attention) exploits the foveal processing structure by directing the high resolution fovea of both image arrays at the same location in space. This fixation process directs the vast majority of the computational resources to a small, well-defined region where the combined resources can resolve the stereo correspondence process in a constrained fashion, and where the (known) vergence angle can be used to assist the computation of scene depths. Thus, for a *given* vergent camera geometry (vergence angle), the computation of depths is simultaneously made *equally difficult (or simple) across the field of view,* and also the perception of depth is made multiscale across the field of view. Foveal perception of depth is therefore detailed and rich in information, while peripheral perception is used largely for contextual processing, peripheral event detection, and coarse-scale model-

building. Along these lines, the goal is to develop paradigms for multiple-channel depth perception based on recent evidence for binocular quadrature-pair receptive fields. The matching of *multiband stereo primitives* (computed, for example, from dual quadrature Gabor arrays) allows for the possibility of partitioning the correspondence problem over multiple channels.

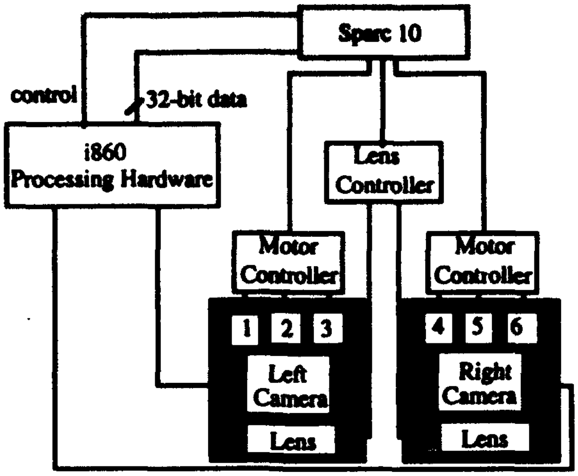
Progress on the Active Vision system has been limited to some degree by the necessity for equipment redesign and construction. Nevertheless, progress has been made both on the hardware and theoretical fronts. One Ph.D. student has been admitted to candidacy on this project.

1. *Design and construction of the Texas Active Vision Testbed (Bovik, Klarquist, Yim).*

Construction is near completion on the redesigned Texas Active Vision Testbed (TAVT). TAVT represents the most recent stage of an ongoing design process that weighed a large number of alternatives for both hardware and software to create a flexible and precise tool for active vision research. Important decisions have been made toward redesign of the original AFOSR-proposed system. A critical aspect of employing active control is providing a means to deal with the large volume of information inherent due to feedback in the acquisition process. A bottleneck in the feedback loop occurs because there is often a different bus structure between the image acquisition and processing components, or between different levels of processing hardware. The TAVT has evolved in this regard. The most recent stage in its evolution is the addition of Dalsa digital cameras to provide two digital data streams to an Alacron dual i860 processor shared memory system for processing and analysis. The advantages of this structure are that the dual i860 processor provides a single processor dedicated to each image, for performing low level processing tasks, while simultaneously allowing memory of the two images to be shared, for stereo processing. The i860 processor provides a balance between the desired speed for low level processing and the flexibility to be programmed for higher level processing such as the evaluation of stereo matching.

2. *Multiresolution, Foveated, Computational Visual Processing for Active Vision (Bovik, Klarquist, Yim).*

Testing and demonstration of TAVT has been extensive. As detailed in an accompanying conference paper [13*] algorithms have been designed and successfully implemented to accomplish variable-baseline stereo bootstrapping and depth-from-focus. Although these algorithms are themselves new, they are fairly simplistic and not based on an assumption of foveated image data. However, we are in the process of adapting the approaches for practical application in an actual foveated processing arrangement that is under development.



3. *Maximum-likelihood focus algorithm (Bovik, Klarquist, Geisler).*

Bovik's lab is developing an algorithm for estimating focus error and object distance in camera systems where the camera focal length can be changed by known amounts. The algorithm is based upon applying a maximum likelihood (ideal-observer) method within a local spatial-frequency analysis. In its current form, the algorithm assumes that there exists an accurate model of the camera's optics, and that both the image and sensor noise are Poisson (or Gaussian).

Consider a camera (or any other optical image-capture system) that has precise adjustable focusing (i.e., a calibrated means of controlling optical focal length). Suppose further that the optical transfer function of the camera is known or has been measured for each possible focus setting (i.e., each focal length). Now consider capturing two (or more) images of the same scene, but with different focal length settings of the camera. (No part of any of the captured images need be in clear focus.) Geisler has derived an optimal (maximum-likelihood) method for estimating the focus error for every (and any) sub-region in the images from the two (or more) images obtained with the camera. The only requirement is that the focal length setting of the camera must be known for each image taken. From the estimates of the focus errors, it is possible to compute the distance from the camera to each point in the image. It is also possible to correct (deblur) the image for the focusing errors that may have occurred at any point in the image. In fact, the maximum-likelihood process provides optimal image restoration from the set of measured images.

Unlike prior approaches, this method has several advantages: (*i*) it allows an accurately measured model of the camera transfer function to be incorporated; (*ii*) it allows for an optimal strategy for depth-from-focus computation, rather than the usual (e.g., heuristics involving edge detector response maximization, etc.); (*iii*) it requires, in theory, only a very few (as few as 2) focusing positions be acquired prior to computing the depth at each pixel; (*iv*) it can be subjected to Ideal Observer methods to determine the actual amount of information available; (*v*) it provides a potentially optimal means of automatic, high-precision focusing; (*vi*) it can be used to simultaneously compute depth maps and deblur images. This

work is very recent. To date, the algorithm has been tested on synthetically blurred images of faces and it appears to work very well. Bovik, Klarquist, and Geisler are currently in the process of implementing the algorithm in Bovik's active vision lab, so that its performance can be evaluated in a functioning computer vision system. At the moment, they are acquiring accurate camera transfer functions to enable testing, and they are further developing a multiband (Gabor channel) framework that will allow for localized processing. Another possible line of investigation will be to compare human accommodation performance with the ideal focusing algorithm.

A technical report presenting the mathematics and a simulation of the maximum-likelihood method is in progress [59]. There is a possibility that some aspects of the algorithm or its implementation in the active vision system will be of commercial value.

C. *Ghosh's Laboratory*

Ghosh and his students have concentrated on two main objectives:

a) To development and analyze mathematical models and algorithms for early visual processing, with emphasis on spatio-temporal processing in neural networks.

b) To study the implementational aspects of the models and algorithms in (a), on workstations, as well as parallel platforms.

Progress on objective (a) is described here; progress on objective (b) is described under Aim 5.

1. *Classification performance of neural networks.*

It has been shown that the outputs of certain trained neural networks approximate Bayesian *a posteriori* probabilities. Thus they might be useful for estimating the information loss (for classification purposes) at various stages of the sequential ideal-observer model developed by Geisler and the UT Vision Group. This possibility prompted a detailed theoretical and experimental study of the classification properties of neural networks, which is reported in [11*]. In a related effort, Ghosh is working with M. Pattichis, a student of Bovik, on the recognition and classification of textured images through feature extraction at multiple resolutions [35].

2. *Clustering of visual patterns.*

Clustering of visual patterns is a task encountered at several levels of visual cognition. A basic issue that needs to be resolved during clustering is the appropriate choice of resolution or scale, which influences the nature and number of clusters. Using an idea from statistical mechanics, a clustering technique has been developed in Ghosh's lab which automatically chooses the appropriate scale(s); it has been applied to medical images, among other patterns

[3*]. The technique uses a biologically plausible neural network structure in an unsupervised, self-organizing fashion [4*]. This effort is being continued because of the promising results, and is expected to appear as a journal article [25].

3. *Spatio-temporal prediction and discrimination*

Ghosh and his students are studying spatio-temporal processing of sequences with emphasis on (i) how to anticipate future inputs in a sequence [18*], and (ii) how to distinguish one sequence from another. For the latter problem, they have developed an Adaptive Spatio-TEmporal Recognizer (ASTER) network that accumulates evidence from previous matches while processing the current input [20]. At present this study is limited to one-dimensional signals (sonar and forecasting problems). The models developed will be applied to images (raw as well as the outputs of the UT vision group's front-end model) in the coming years.

C. *Super's Laboratory*

The development of computer-vision algorithms for performing many complex tasks requires a thorough understanding of the relationship between surfaces, lighting, and the images produced in the camera or eye. The broad objectives of Super's research are:

(a) To further develop and extend his Surface-to-Image-Projection (STIP) model.

(b) To develop algorithms to perform visual tasks by exploiting the STIP model.

There is little doubt that biological vision systems have evolved to exploit surface-to-image projection constraints; thus the research in Super's laboratory is playing an important role in guiding the group's development of both computer vision models and models of human perception.

To date, Super's research has concentrated on the detailed development of the STIP model, its use in computing surface shape and orientation from texture cues, and the measurement of local spatial frequency information in the image for this purpose. Now, the emphasis of the research is on exploring how to combine multiple sources of information to recover 3-D geometry. Of particular interest is the use of the STIP model with multiple views (stereo and motion), and with shading and shadow information. A newly evolving direction for the research is use of the multi-view STIP model in the active vision testbed. The STIP model correctly captures the geometrical distortions of image patches between views; Super is starting to explore its use to make stereo matching more reliable, and to control vergence [with Klarquist, Balasubramanyam].

In addition, exploratory studies are underway on the use of the outputs of local spatial frequency channels for perceptual grouping and for matching. Efforts are also being directed at developing techniques for identifying structure in very noisy data [with Geisler].

1. *STIP for perspective projection (Super).*

The STIP model has been extended to describe the projection of local spatial and local spatial frequency quantities under perspective projection. [16*, 17*]

2. *Image filters for direct detection of surface orientation (Super).*

Super has further developed the use of the STIP model to define a set of image filters for directly detecting surface orientation from texture information. In this approach, rather than decompose the image texture into sinusoidal components, the image texture is decomposed into basis elements that are variable-frequency image sinusoidal gratings, whose structure reflects perspective deformations directly. Current research is testing a binocular version of these filters and applying them to non-texture stimuli. [37]

3. *STIP for binocular (two-view) vision (Super, Chen).*

The STIP model has been extended to the two-view case. Super is exploring binocular versions of the filters in project (2) above [37], and in other work, the use of shape from texture in two views to compute the inter-view rotation matrix. The stereo work is not limited to parallel baseline stereo but is completely general; thus there is no distinction between stereo and frame-based motion in this work. In addition, the STIP model has been used for stereo matching based on phase outputs of banks of Gabor wavelet filters. [6*]

4. *Contrast normalization in shape from texture (Super).*

The effect of incorporating non-linear contrast normalization into the shape-from-texture algorithms that use the STIP models was examined. Comparison of the contrast-normalization mechanisms used by Super and those found in the cortex by Albrecht and Geisler shows that they have similar effects when applied to texture images. However, Albrecht and Geisler's physiological model incorporates terms (for example, to account for the finite rise-time of neuronal activity) that Super's does not; Super's is a simple version suitable for image processing. Interestingly, contrast normalization provides a powerful method for separating texture information from the shading information in images. The technique is used for this purpose in [40*, 16*, 17*]. It remains to be seen whether the human visual system uses contrast normalization for similar purposes.

5. *AM-FM image models in shape from texture (Super).*

The more accurate local spatial frequency estimation techniques developed by Bovik et al. for AM-FM image models have been incorporated into Super's shape-from-texture algorithms, resulting in improved performance of the latter [16*, 17*, 37].

IV. Aim 4: *Develop models for human performance in complex visual tasks that build upon current understanding of the front-end mechanisms.*

A. *Cormack's Laboratory*

1. *Coarse-to-fine processing in stereoscopic vision (Cormack, Chen, Ramakrishnan).*

This study (in which pilot data are currently being collected) is being done by Cormack in collaboration with Chen and Ramakrishnan, both of the Department of Computer Engineering. The study concerns the manner in which the visual system combines the phase information (contained within the different spatial-frequency bands of the two monocular images) in order to achieve the large disparity range and fine disparity resolution that the visual system possesses. Cormack et al. wish to determine if the presence of low spatial frequency information can resolve the depth ambiguity inherent in the cyclic nature of phase disparities for high spatial frequencies. This type of processing is referred to as a "coarse-to-fine processing strategy" and is being employed in a computational model of stereoscopic processing developed by Chen. Based on the outcome of this and future studies, Ccrmack and Ramakrishan will attempt to incorporate the type of processing employed by the human visual system into biologically plausible models of stereopsis. This study will also serve as groundwork for future collaborative studies between Chen and Cormack.

B. *Geisler's Laboratory*

1. *Measurements and models of visual search (Geisler, Chou, Kortum).*

Visual search, under conditions that require multiple fixations, is a complex but fundamental task that is strongly affected by both low-level factors (such as the information content of the stimuli and the parallel and automatic mechanisms at the front end of the visual system), as well as by high-level factors (such as attention mechanisms, eye-movement control mechanisms and decision processes). In order to rigorously study human visual search performance, Geisler and Chou have developed a theoretical approach and an experimental method for assessing (and hence isolating) the role of low-level factors in complex tasks. The method involves comparing simple-discrimination performance and visual search performance for the same stimuli. In one of the completed experiments, the target and background were composed of line segments that differed in color and/or orientation; in another experiment, the target and background were composed of filtered-noise textures that differed in spatial frequency and/or orientation. Analysis of the results showed that much of the variance in search time was predictable from the discrimination data, suggesting that low-level factors often play a dominant role in limiting search performance. Geisler and Chou also developed a signal-detection model which demonstrates how current psychophysical models of visual discrimination might be generalized in order to obtain a quantitative theory that can predict visual search performance under a wide range of stimulus conditions. The results of this study will be published in *Psychological Review* [10*].

There were two weaknesses in the Geisler and Chou study. First, the model of visual search is too simple (serial, non-overlapping fixations). Geisler is currently working on a more realistic model which includes a mechanism for double checking during search. Second, the experiments did not involve measuring eye movements. Kortum and Geisler are currently modifying the experiment so that eye movements can be recorded during both the discrimination and search tasks. The plan is to use actually recorded eye movements as inputs to the model. This will provide a much stronger test of the model (and undoubtedly motivate further changes in the model).

2. ***Finding the spatial structure in 2-D images (Geisler, Super).***

A key issue in visual science is how the local spatial information extracted by the front-end of the visual system is combined in order to perform complex visual tasks. This study (which is still in its beginning stages) concerns how such local measurements are combined to find the spatial structure that exists in 2-D images (an ability that is very well-developed in the human visual system). Many key insights into the mechanisms that the human visual system employs were provided by the demonstrations (of grouping and segregation processes) devised by the Gestalt psychologists and by many perception researchers that have followed. The strategy that Geisler and Super are taking in attacking this difficult problem is to begin by developing a simple computational model that operates on a restricted stimulus domain, yet incorporates most of the known grouping rules within a single coherent framework. The stimulus domain that they are considering is the set of images that can be defined by a relatively small list of coordinate pairs, e.g., a small collection of points, lines or polygons. To be concrete, the reader might think of the input as a list of oriented line segments defined by their endpoints (plus their gray level and color); such a list might approximately describe what is extracted by the front end of a visual system. Geisler and Super have developed a minimum-squared-error model in which collections of coordinate pairs (e.g., line segments) are grouped using a weighted combination of shape, orientation, distance, symmetry, size (scale), gray-level, and color. (This model has been implemented in a C program which takes lists of coordinate points as input.) One question Geisler and Super are asking is whether such a weighted sum of grouping rules/dimensions can predict the structure that humans "see" in images created within the defined stimulus domain. So far, the experiments have only involved viewing images and then comparing subjective judgments with the output of the model. The model is able to capture a number of image structures that humans report subjectively. Currently, efforts are being directed (a) at developing a quantitative information metric of the amount of redundancy removed from the image by the grouping processes, and (b) at finding a rigorous experimental paradigm that will allow estimation of the weights that subjects place on the different grouping rules/dimensions.

C. *Gilden's Laboratory*

1. *Noise processes in spatial and temporal interval estimation (Gilden).*

It is typically supposed that the noise (errors) in perceptual estimation is normally distributed (the common assumption in statistical tests of significance); however, there have been almost no efforts to actually measure perceptual estimation noise. Gilden has discovered that the Fourier spectrum of the estimation noise, when estimating spatial and temporal intervals, falls off with a slope of -1 in log-log coordinates (1/f noise). This is in sharp distinction to the expectations from a normal distribution (Gaussian noise) which should have a slope of 0.0. Ten experiments have been run, and the domain in which 1/f noise arises has been mapped out. A paper on this research has been submitted to *Science* [21*]. Theoretical mechanisms that produce 1/f noise are being studied with respect to their implementation in neural networks. These findings may have a wide impact upon the development of models for perceptual estimation in many stimulus domains.

2. *The role of frames of reference in motion processing (Gilden)*

Gilden has developed a theory of motion perception based on mathematical and visual constraints on the formation of frames of reference. He is attempting to distinguish between frames of reference as conceived mathematically, and perceptual frames of reference. From a mathematical point of view, any motion field can locally serve as a coordinate system for defining the positions and velocities of other motions. However, there are attentional constraints in visual analysis that permit only translation fields to serve as frames of reference. A theorem has been constructed that articulates these concepts. Its gist is that if and only if a motion field can be fully represented by diagonal energy in space-time can it be processed preattentively. Only translation fields satisfy this requirement. Rotation fields, for example, require spatial distinctions with respect to the axis of rotation, in addition to the specification of local speed. That is, clockwise rotation has the upper part of the field moving right and the lower part of the field moving left. For homogeneous translation fields there is no local specification of spatial concepts such as upper and lower. The content of the theorem is that such spatial references are consuming of attentional resources. A corollary of this theorem is that only translation fields can serve as global reference frames. This follows from the fact that all other motion fields (rotation, divergence, and shear) specify geometric elements and the pinning down of these elements in space exhausts attention. If so, then multiple rotation (shear, divergence) fields must be analyzed serially. For one motion field to serve as a frame of reference for a second motion field, it is necessary that both be processed simultaneously. Gilden is attempting to develop the implications of this theory for memory, attention, perceptual organization, and reasoning. Fourteen experiments on the encoding of motion fields in memory have been conducted. Three experiments are now being planned for assessing how fields with multiple rotations are processed. A paper is in preparation that will be submitted to *Psychological Review* [31].

V. Aim 5: ***Develop a computational testbed for implementing, comparing, integrating and visualizing the different models and modules developed during the project, using a massively parallel machine and graphics workstation front-end.***

A. *Ghosh's Laboratory*

Early in the project, it was decided to move the purchase of the MasPar to the second year. With hindsight, this has been a sound decision, as we were able to obtain a machine with almost 4 times the power for about the same price. The 4K processor MasPar MP-1 was installed on June 27, 1994.

Since the MasPar machine purchase was postponed by a year, Ghosh used this time to study the algorithmic demands (mapping, processing, memory, I/O) of several algorithms and models developed by Bovik's and Geisler's groups. In particular, Ghosh's group has implemented early vision simulation software using Matlab, in close cooperation with Geisler. This simulator models the optics, receptors, ganglion cells and frequency selective cortical cells, and incorporates the effect of varying resolution/cell density as a function of eccentricity. The effects of eccentricity on discrimination of textures at different levels of the model have been experimentally studied using this simulator. The classification performance drop with increasing eccentricity and with increasing noise is being examined. We have found the most noticeable drops at the outputs of ganglion cells, and have also encountered certain situations where aliasing effects actually improve the performance [33].

It is anticipated that two key issues in implementing the Vision Group's models on the MasPar will be (i) effective mapping and (ii) input/output (I/O). Ghosh's lab has initiated analysis of I/O traffic at both the interconnection level [36] and external disk level [30], and is also studying mapping and load balancing techniques for multiresolution algorithms/data, such as processor allocation in the presence of subsampling and eccentricity. Finally, substantial effort has been put to seeing that the various members of the UT vision group have common working environments (Khoros, Eudora etc.), since this will facilitate continuing interactions.

Publications:
Publications marked * are appended.

1993

1.* Cormack, L. K., Stevenson, S. B. & Schor, C. M. (1993) Disparity-tuned channels of the human visual system. *Visual Neuroscience, 10,* 585-596.

1994

2.* Albrecht, D. G. and Geisler, W. S. (1994) Visual cortex neurons in monkey and cat: Contrast response nonlinearities and stimulus selectivity. In: T. Lawton (Ed.), Computational Vision Based On Neurobiology, *SPIE Proceedings*, *2054*, 12-31.

3.* Chakravarthy, S. V. and Ghosh, J. (1994) Scale-based clustering using a radial basis function network. *Proceedings of the IEEE Int'l Conf. on Neural Networks*, 897-902.

4.* Ghosh, J. and Chakravarthy, S. V. (1994) The rapid kernel classifier: A link between the self-organizing feature map and the radial basis function network. *Jl. of Intelligent Material Systems and Structures, (special issue on neural networks)*, *5*, 211-219.

5.* Stevenson, S. B., Cormack, L. K. & Schor, C. M. (1994) The effect of stimulus contrast and interocular correlation on disparity vergence. *Vision Research*, *34*, 383-396.

in press

6.* Chen, T.-Y. , Bovik, A. C. , and Super, B. J. Multiscale stereopsis via Gabor filter phase response. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, in press.*

7.* Cormack, L. K., Stevenson, S. B. & Schor, C. M. An upper limit to the binocular combination of stimuli. *Vision Research, in press.*

8.* Geisler, W. S. Discrimination information in natural radiance spectra. In: E. Peli (Ed.), *Vision Models for Target Detection and Recognition.* World Scientific Publishing, Co, *in press.*

9.* Geisler, W. S. and Banks, M. S. Visual performance. In: *Handbook of Optics.* New York: McGraw-Hill, *in press.*

10.* Geisler, W. S. and Chou, K. L. Separation of low-level and high-level factors in complex tasks: Visual search. *Psychological Review, in press.*

11.* Ghosh, J. and Tumer, K. Structural adaptation and generalization in supervised feedforward networks. *Jl. of Artificial Neural Networks, in press.*

12.* Hahn, L. and Geisler, W. S. Adaptation mechanisms in spatial vision I: Bleaches and backgrounds. *Vision Research, in press.*

13.* Klarquist, W. N. and Bovik, A. C. The Texas active vision testbed. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.*

14.* Kortum, P. T. and Geisler, W. S. (1994) Adaptation mechanisms in spatial vision II: Flash thresholds and background adaptation. *Vision Research, in press.*

15.* Shin, Y. and Ghosh, J. Ridge polynomial networks. *IEEE Transactions on Neural Networks, in press.*

16.* Super, B. J. and Bovik, A. C. Planar surface orientation from texture spatial frequencies. *Pattern Recognition, in press.*

17.* Super, B. J. and Bovik, A. C. Shape from texture using local spectral moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence, in press.*

18.* Tumer, K. and Ghosh, J. Sequence recognition by input anticipation. *Proceedings of the IEA/AIE'94, in press.*

19.* Yim, C., Klarquist, W. N. and Bovik, A. C. Multiresolution feature extraction based on multifrequency decomposition. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, in press.*

under review

20. Ghosh, J. and Deuser, L. Classification of spatio-temporal patterns with applications to recognition of sonar sequences. *Neural Representation of Temporal Patterns*, E. Covey, H. Hawkins, T. McMullen and R. Port (Eds.), *under review.*

21.* Gilden, D. L, Thornton, T., & Mallon, M. 1/f noise in human cognition. *Science, under review.*

in preparation

22. Albrecht, D. G. Visual cortex neurons in monkeys and cats: The effects of contrast on the phase transfer function. *Visual Neuroscience, in preparation.* Expected submission date: October, 1994.

23. Albrecht, D. G. and Geisler, W. S. The effects of contrast nonlinearities and response variability on detection, discrimination and identification performance of visual cortex neurons. *Vision Research, in preparation.* Expected submission date: December, 1994.

24. Bovik, A. C., Desai, M. D., and Havlicek, J. P. Limits on discrete modulated signals. *IEEE Transactions on Signal Processing, in preparation.*

25. Chakravarthy, S. V. and Ghosh, J. Scale-based clustering with unsupervised RBF networks. *In preparation. Journal to be determined.*

26. Cormack, L. K. & Landers, D. L. Element density and the efficiency of binocular matching. *Vision Research, in preparation.* Expected submission date: January, 1995.

27. Cormack, L. K., Landers, D. L. & Mallon, M. W. A crossed/uncrossed disparity anisotropy for the speed of stereoscopic processing. *Perception and Psychophysics, in preparation.* Expected submission date: January, 1995.

28.* Cormack, L. K. & Riddle, R. Binocular correlation using oriented contrast information. *Vision Research*, in *preparation.* Expected submission date: August, 1994.

29.* Geisler, W. S. and Albrecht, D. G. Identification performance in the primate visual system: Selectivity and the certainty function. *Science, in preparation.* Expected submission date: August, 1994.

30. Ghosh, J. and Gupta, S. Novel disk drive designs for massively parallel vision applications. *In preparation. Journal to be determined.*

31. Gilden, D. L., & Proffitt, D. R. Perceptual constraints on the construction of frames of reference. *Psychological Review*, in *preparation.* Expected submission date: September, 1994.

32. Havlicek, J. P. and Bovik, A. C. Multi-component amplitude and frequency modulation models for image processing and computational vision. *IEEE Transactions on Image Processing, in preparation.*

33. Kuyel, T., Geisler, W. S. and Ghosh, J. The effects of eccentricity in texture discrimination performance using a simulated biological early vision system. *In preparation.*

34. Maragos, P. and Bovik, A. C. Demodulation of images modeled by amplitude-frequency modulations using multidimensional energy

separation. *Journal of the Optical Society of America, in preparation*.

35. Pattichis, M., Garcia, I., Ghosh, J. and Bovik, A. C. Multiresolution processing of textured images. *In preparation.*

36. Ramamurti, V. and Ghosh, J. Frequency analysis of I/O traffic in massively parallel systems with applications to early vision. *In preparation.*

37. Super, B. J. Direct detection of 3D surface orientation with monocular and binocular image filters. *In preparation. Journal to be determined.* Expected submission date: September, 1994.

Conference Presentations/Talks/Technical Reports:

1993

38.* Cormack, L. K. (May, 1993). *Annual Meeting of the Association for Research in Vision and Ophthamology.* Interocular correlation detection with oriented contours.

39.* Gilden, D. L., Hiris, E., and Blake, R. (May, 1993). *Annual Meeting of the Association for Research in Vision and Ophthamology.* Motion coherence is perceived as information entropy.

40.* Super, B. J., Bovik, A. C., and Geisler, W. S. (May, 1993). *Annual meeting of the Association for Research in Vision and Opthalmology.* A model of shape from texture using second-order moments of local spatial-frequency spectra.

41. Geisler, W. S. (July, 1993). *University of Minnesota.* Low level and high level factors in visual performance.

42. Geisler, W. S. (August 17, 1993). *U. S. Army Research Labs, Aberdeen Proving Ground, MD.* Sequential analysis of visual performance.

43. Cormack, L. K. (September, 1993). *University of Houston.* Binocular matching and stereopsis.

44. Geisler, W. S. (November 16, 1993). *The Rockefeller University.* A model of visual discrimination performance.

1994

45. Geisler, W. S. (March, 1994). *School of Optometry, University of Houston.* Model of single neuron responses in primary visual cortex.

46. **Chen, T.-Y., Klarquist, W. N. and Bovik, A. C. (April, 1994). *IEEE Southwest Symposium on Image Analysis and Interpretation.* Stereo vision using Gabor wavelets.**

47. Geisler, W. S. (April, 1994). *University of Texas Health Sciences Center, Houston.* Optical and retinal factors in spatial vision and adaptation.

48. Yim, C. and Bovik, A. C. (April, 1994). *IEEE Southwest Symposium on Image Analysis and Interpretation.* Vergence control using hierarchical image structure.

49.* Balasubramanyam, P. (May, 1994). *Annual meeting of the Association for Research in Vision and Ophthalmology.* Motion perception from local motion and disparity measurements.

50.* Cormack, L. K. and Landers, D. D. (May, 1994). *Annual Meeting of the Association for Research in Vision and Ophthamology.* The efficiency of binocular matching.

51.* Geisler, W. S. (May, 1994). *Annual meeting of the Association for Research in Vision and Ophthalmology.* Optical and retinal factors in spatial vision and adaptation.

52.* Gilden, D. L. (May, 1994). *Annual meeting of the Association for Research in Vision and Ophthalmology.* Streaky performance in visual psychophysics.

53.* Hahn, L. W. and Geisler, W. S. (May, 1994). *Annual meeting of the Association for Research in Vision and Ophthalmology.* Local light-adaptation mechanisms in grating adaptation.

54.* Kortum, P.T. and Geisler, W. S. (May, 1994). *Annual meeting of the Association for Research in Vision and Ophthalmology.* Light adaptation mechanisms in spatial target detection.

55. Ghosh, J. (June, 1994). *IEEE Int'l Conf. on Neural Networks.* Scale-based clustering using a radial basis function network.

56. Tumer, K. (June, 1994). *IEA/AIE.* Sequence recognition by input anticipation.

anticipated

57. Chen, T.-Y. , Bovik, A.C. and Super, B. J. (October, 1994). *IEEE International Conference on Systems, Man, and Cybernetics.* Multiscale stereopsis via Gabor filter phase response.

58. Geisler, W. S. (August, 1994). *Annual Meeting of the Society of Mathematical Psychology.* Multiple fixation search: Mechanisms and models.

59. Geisler, W. S. and Klarquist, W. A maximum likelihood method for computing focus error, distance, and focus correction in camera images. *CVIS TR-94005.*

60. Chen, T.-Y., Bovik, A. C. and Super, B. J. (October, 1994). *IEEE International Conference on Systems, Man, and Cybernetics.* Multiscale stereopsis via Gabor filter phase response.

61. Klarquist, W. N. and Bovik, A. C. (October, 1994). *IEEE International Conference on Systems, Man, and Cybernetics.* The Texas active vision testbed.

62. Yim, C., Klarquist, W. N. and Bovik, A. C. (October, 1994). *IEEE International Conference on Systems, Man, and Cybernetics.* Multiresolution feature extraction based on multifrequency decomposition.

63. Maragos, P. and Bovik, A. C. (November, 1994). *IEEE International Conference on Image Processing.* Demodulation of images modeled by amplitude-frequency modulations using multidimensional energy separation.

64. Havlicek, J. P. and Bovik, A. C. (November, 1994). *IEEE International Conference on Image Processing.* Multi-component AM-FM image models and wavelet-based demodulation with component tracking.

CVIS Personnel, grant year I:

Principal Investigators:
Wilson Geisler
Alan Bovik
Joydeep Ghosh
David Gilden
Lawrence Cormack
Boaz Super

Technical Support Personnel:
Carl Creeger
Larry Stern

Administrative Support Personnel:

Kirsten Bradbury	(March, 1994-present)
Constance Quigley	(January, 1994-March, 1994)
Susan Fawver	(May, 1993-January, 1994)

Direct Research Personnel:

Graduate Students:
K. Bollacker
S. Chakravarty
Tieh-Yuh (Terry) Chen
Kgay-Lay Chou
S. Gupta
Lance Hahn
Joseph Havlicek
William Klarquist
Phil Kortum
Turker Kuyel
Marios Pattichis
Christy Price
Siva Ramakrishnan
V. Ramamurti
Christina Seay
Kagan Tumer
Chagoon Yim

Undergraduate Students:
Young Mi Kim
David Landers

Non-Students:
Poornima Balasubramanyam
Elizabeth Robinson